

Doc Code: AP PRE REQ

PTO/SB/33 (07-05)

Approved for use through xx/xx/200x. OMB 0551-00xx
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PRE-APPEAL BRIEF REQUEST FOR REVIEW		Docket Number (Optional)	
		YOR920030413US1	
I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to "Mail Stop AF, Commissioner for Patents P O Box 1450 Alexandria VA 22313-1450" [37 CFR 1.8(a)] on _____ Signature _____ Typed or printed name _____	Application Number	Filed	
	10/722,776	November 26, 2003	
	First Named Inventor		
	Haixun Wang		
	Art Unit	Examiner	
	2624	Nancy Bitar	
Applicant requests review of the final rejection in the above-identified application. No amendments are being filed with this request.			
This request is being filed with a notice of appeal.			
The review is requested for the reason(s) stated on the attached sheet(s). Note: No more than five (5) pages may be provided.			
I am the			
<input type="checkbox"/>	applicant/inventor	/Kevin M. Mason/ Signature	
<input type="checkbox"/>	assignee of record of the entire interest. See 37 CFR 3.71. Statement under 37 CFR 3.73(b) is enclosed (Form PTO/SB/93)	Kevin M. Mason Typed or printed name	
<input checked="" type="checkbox"/>	attorney or agent of record Registration number 36,597	(203) 255-6560 Telephone number	
<input type="checkbox"/>	attorney or agent acting under 37 CFR 1.34 Registration number if acting under 37 CFR 1.34 _____	February 19, 2008 Date	
NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below.			
<input type="checkbox"/> *Total of _____ forms are submitted.			

This collection of information is required by 35 U.S.C. 132. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11, 1.14 and 41.6. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Mail Stop AF, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

If you need assistance in completing the form call 1-800-PTO 9199 and select option 2

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

5 Applicant(s): Wang et al.
Docket No.: YOR920030413US1
Serial No.: 10/722,776
Filing Date: November 26, 2003
Group: 2624
10 Examiner: Nancy Bitar

Title: Near-Neighbor Search in Pattern Distance Spaces

15

MEMORANDUM IN SUPPORT OF
PRE-APPEAL BRIEF REQUEST FOR REVIEW

20 Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

25

Sir:

The present invention and prior art have been summarized in Applicants' prior responses.

30

STATEMENT OF GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The present application was filed on November 26, 2003 with claims 1 through 20. Claims 1 through 20 are presently pending in the above-identified patent application. Claims 1-20 were rejected under 35 U.S.C. §102(b) as being anticipated by Wang et al. ("Clustering by Pattern Similarity in Large Data Sets," ACM SIGMOD 2002
35 June 4-6, Madison, Wisconsin, USA)

ARGUMENTS

Independent Claims 1 and 18-20

40 Independent claims 1 and 18-20 were rejected under 35 U.S.C. §102(b) as being anticipated by Wang et al. Regarding claim 1, the Examiner asserts that Wang

teaches identifying subspace pattern similarities that the objects in the set exhibit in multi-dimensional spaces (section 1.3); and defining subspace correlations between two or more of the objects in the set based on the identified subspace pattern similarities for use in identifying near-neighbor objects. The Examiner asserts that Wang discloses clustering by pattern similarity. In the Response to Arguments section of the final Office Action, the Examiner asserts that when each object represents its own cluster, the distances between those objects are defined by the chosen distance measure, and that once several objects have been linked together, the distances between those new clusters are determined by finding the "nearest neighbors" across clusters to determine the distances between clusters. The Examiner asserts that therefore Wang teaches the method where the similarity model used in data retrieval and nearest neighbor search is based on value similarity (section 6).

As the Examiner acknowledges, Wang is directed to clustering by pattern similarity. (See, Abstract.) While the processes of "clustering" and "finding the nearest neighbor" share the concept of pattern similarity, the results of the processes are *not* the same, as would be apparent to a person of ordinary skill in the art. For example, in "clustering," a given set of datasets are processed and grouped into clusters. Once the clustering is completed, however, the nearest neighbor of a given data item is still not known. Thus, "clustering" and the Wang reference do *not* disclose or suggest defining subspace correlations between two or more of the objects in the set based on the identified subspace pattern similarities for use in identifying near-neighbor objects.

Moreover, the "Finding Nearest Neighbor" technique and "Clustering" technique are very different statistical analysis tools. While most of the published and patented work in these two art forms use the same distance function (or similarity metric, such as the Euclidean distance), a person of ordinary skill in the art would understand that the "Nearest Neighbor" techniques cannot be derived from the "Clustering" work simply based on the fact that the similarity measure is the same.

Specifically, the Examiner has asserted that "... since when each object represents its own cluster; the distances between those objects are defined by the chosen distance measure. However, once several objects have been linked together, we

determine the distances between those new clusters by finding the 'nearest neighbors' across clusters to determine the distances between clusters ...".

Applicants respectfully disagree

First, if the case where each object is its own cluster is considered, *then no*
5 *clustering has been performed at all*. In this case, the purpose of nearest neighbor search is to find the nearest neighbor in the raw data, with no assistance from clustering whatsoever.

Second, if clusters have been discovered, the search for the nearest neighbor will not be easier, as would be apparent to a person of ordinary skill in the art
10 The quality of clustering is measured by a global goal, which is the purity within clusters, and dissimilarity between clusters. The quality function is a value aggregated over the entire data set. Thus, it is easy to conceive a case where the nearest neighbor of an object in cluster A is in cluster B, because these two objects have the shortest distance. If you move the two objects into one cluster, however, the global quality of clustering will be
15 reduced because they need to interact with the other objects in the clusters. In the present case, this phenomenon is even more obvious because subspace clustering is being considered, which means the two clusters mentioned above are very likely clusters in different subspaces, which essentially means that the distance functions they use to evaluate similarity are totally different, and certainly very likely different from the
20 distance function used to measure the similarity between the two particular objects. Thus, clustering will not help the process of finding the nearest neighbor at all.

Thus, as noted above, "clustering" and the Wang reference do *not* disclose or suggest defining subspace correlations between two or more of the objects in the set based on the identified subspace pattern similarities for use in identifying near-neighbor
25 objects. Independent claims 1, 19, and 20 require identifying subspace pattern similarities that the objects in the set exhibit in multi-dimensional spaces; and defining subspace correlations between two or more of the objects in the set based on the identified subspace pattern similarities for use in identifying near-neighbor objects. Independent claim 18 requires defining subspace correlations between two or more of the
30 objects in the set based on the identified subspace pattern similarities; and using the

subspace correlations to identify near-neighbor objects among the query objects and the objects in the set.

Thus, Wang et al. do not disclose or suggest identifying subspace pattern similarities that the objects in the set exhibit in multi-dimensional spaces; and defining
5 subspace correlations between two or more of the objects in the set based on the identified subspace pattern similarities for use in identifying near-neighbor objects, as required by independent claims 1, 19, and 20, and do not disclose or suggest creating a pattern distance index to identify subspace pattern similarities that the objects in the set exhibit in multi-dimensional spaces; defining subspace correlations between two or more
10 of the objects in the set based on the identified subspace pattern similarities; and using the subspace correlations to identify near-neighbor objects among the query objects and the objects in the set, as required by independent claim 18.

Dependent Claims 2-17

Dependent claims 2-17 were rejected under 35 U.S.C. §102(b) as being
15 anticipated by Wang et al.

Claims 2-17 are dependent on claim 1 and are therefore patentably distinguished over Wang et al. because of their dependency from independent claim 1 for the reasons set forth above, as well as other elements these claims add in combination to their base claim.

20 All of the pending claims, i.e., claims 1-20, are in condition for allowance and such favorable action is earnestly solicited.

If any outstanding issues remain, or if the Examiner has any further suggestions for expediting allowance of this application, the Examiner is invited to contact the undersigned at the telephone number indicated below.

The Examiner's attention to this matter is appreciated.

Respectfully submitted,

/ Kevin M. Mason/

5

Date: February 19, 2008

10

Kevin M. Mason
Attorney for Applicants
Reg. No. 36,597
Ryan, Mason & Lewis, LLP
1300 Post Road, Suite 205
Fairfield, CT 06824
(203) 255-6560